Artificial Intelligence

Chapter 1

Outline

- What is AI?
- A brief history
- The state of the art

What is AI?

AI is the study and construction of systems that exhibit intelligent behavior, where intelligence can be understood as:

Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

Acting humanly
Building machines whose behavior is indistinguishable from that of humans (e.g., passing the Turing Test).

Thinking humanly
Modeling the actual cognitive processes humans use when they think and reason (cognitive modeling).

Thinking rationally
Using formal logic and correct reasoning ("laws of thought") to reach valid conclusions.

Acting rationally (the book's preferred view)
Designing rational agents—systems that perceive their environment and act to achieve the best expected outcome, given their knowledge and percepts.

What is AI?

The study and construction of agents that do the right thing.

What "the right thing" means is defined by an objective or performance measure supplied to the agent

AI is not just machine learning (ML is a subfield of AI) or Deep Learning, or LLMs.

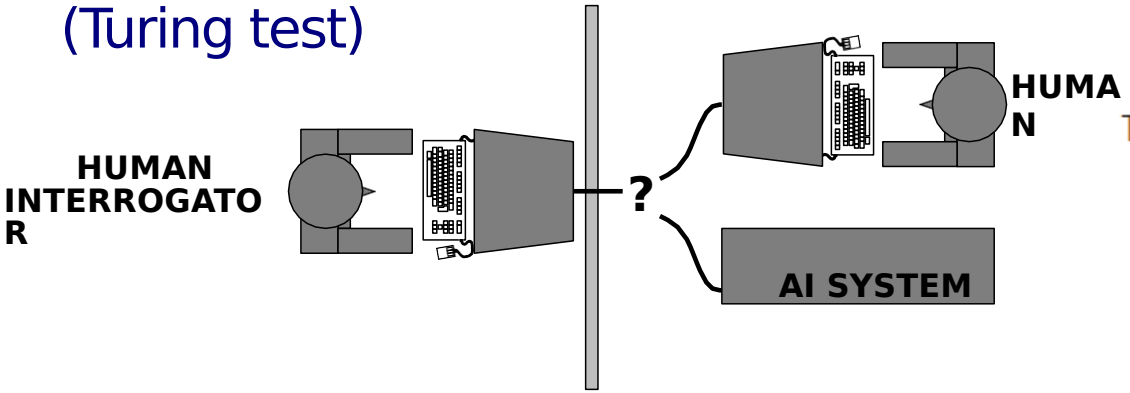
The focus of the class is AI as rational agents acting under uncertainty.

Intelligence is judged by behavior and outcomes, not by whether the system thinks exactly like a human.

Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

- "Can machines think?" —→ "Can machines behave intelligently?"
- Operational test for intelligent behavior: the Imitation Game (Turing test)



Turing argued that this question is **too ambiguous**:

- What does *think* mean?
- How do we measure it?
- Who decides?

- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problem: Turing test is not reproducible, constructive, or amenable to mathematical analysis

© 2021 Pearson Education Ltd.

Turing test

A machine that passes the Turing Test must integrate several AI subfields:

Natural Language Processing: communicate fluently
Knowledge Representation: store and organize facts
Automated Reasoning: answer questions, draw conclusions
Machine Learning: adapt and improve over time

In the Total Turing Test, additional abilities are required:
Computer Vision
Robotics

AI moved beyond the Turing Test

Thinking humanly: Cognitive Science

Before the 1960s, psychology was dominated by behaviorism: Behaviorism studied only observable behavior.

- Internal mental states were considered unscientific. The brain was treated as a black box

In the 1960s, this view was overturned by the cognitive revolution: Humans began to be modeled as information-processing systems.

- Mental states such as beliefs, goals, and memories became legitimate objects of study

The new paradigm viewed the mind as:

- Receiving inputs
- Internally representing information
- Performing transformations (reasoning, memory, planning)
- Producing outputs (actions, speech)

Requirement: theories of internal brain activity

Unlike behaviorism, **thinking humanly** requires:

- Scientific theories of **internal mental processes**
- Explicit models of *how* reasoning, memory, and perception work

Thinking humanly: Cognitive Science

Requires scientific theories of internal activities of the brain

- What level of abstraction? "Knowledge" or "circuits"?
- How to validate? Requires
  1) Predicting and testing behavior of human subjects (top-down) or 2) Direct identification from neurological data (bottom-up)

Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI

Both share with AI the following characteristic:

the available theories do not explain (or engender) anything resembling human-level general intelligence

Hence, all three fields share one principal direction!

© 2021 Pearson Education Ltd.

Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes? Several Greek schools developed various forms of logic:

notation and rules of derivation for thoughts;

may or may not have proceeded to the idea of mechanization

Direct line through mathematics and philosophy to modern AI Problems:

1) Not all intelligent behavior is mediated by logical deliberation
2) What is the purpose of thinking? What thoughts should I have out of all the thoughts (logical or otherwise) that I could have?

© 2021 Pearson Education Ltd.

Thinking rationally: Laws of Thought

Define intelligence as **thinking correctly**, according to formal rules of reasoning.
How should one think or act if one is rational? "The laws of thought" are normative

Normative (Prescriptive): How *should* one think or act if one is rational?

- Defines ideal or correct behavior
- Specifies rules, standards, or principles
- Independent of how humans actually behave
- Concerned with optimality and correctness

Descriptive: How do humans actually think or behave

- Empirical and observational
- Based on experiments and data
- Describes real behavior, including errors and biases
- No claim that the behavior is optimal or correct

Thinking rationally: Laws of Thought
Normative vs Descriptive Approach

Aspect	Normative	Descriptive
Question	What *should* be done?	What *is* done?
Goal	Correctness / optimality	Accuracy of description
Method	Logic, math, optimization	Experiments, observation
Human errors	Ignored	Central
Example	"This is the rational choice"	"People often choose otherwise"

© 2021 Pearson Education Ltd.

Acting rationally

This is the **dominant modern definition of AI**: intelligence as **rational behavior**, not just thought.

Rational behavior: doing the right thing

The right thing: that which is expected to maximize goal achievement, given the available information

Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

> Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good

© 2021 Pearson Education Ltd.

Acting rationally

1. Rational behavior: “doing the right thing.”

In this view, an agent is intelligent if it acts appropriately in its environment.
What matters is what the agent does, not *how* it does it.

2. What is “the right thing”?

That which is expected to maximize goal achievement, given the available information

Expected:
The agent reasons under uncertainty; outcomes are probabilistic.

Maximize:
The agent chooses the best action among alternatives.

Goal achievement:
Behavior is evaluated relative to explicit objectives or performance measures.

Given the available information:
Rationality depends on what the agent knows, not on omniscience.

Rationality is about **expected performance**, not guaranteed success.

Acting rationally - **Why this matters for AI**

This approach succeeds where earlier ones struggled:

It handles **uncertainty**

It accommodates **learning and reflexes**

It supports **mathematical analysis**

It applies to humans, animals, and machines alike

Rational agents

An agent is an entity that perceives (sensors) and acts (actuators)

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : P^* \rightarrow A$$

Where:

= the set of all possible percept sequences (entire history)
= the set of possible actions
= the agent function

Given everything the agent has perceived so far, it chooses an action.

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Caveat: computational limitations

© 2021 Pearson Education Ltd.

AI prehistory

Philosophy	logic, methods of reasoning mind as physical system
Mathematics	foundations of learning, language, rationality
Psychology	formal representation and proof
Economics	algorithms, computation, (un)decidability, (in)tractability probability
Linguistics	adaptation
Neuroscience	phenomena of perception and motor control experimental techniques (psychophysics, etc.)
Control theory	homeostatic systems, stability
	formal theory of rational decisions knowledge representation grammar
	simple optimal agent designs
	plastic physical substrate for mental activity

© 2021 Pearson Education Ltd.

Potted history of AI

1943 McCulloch & Pitts: Boolean circuit model of brain

1950 Turing's "Computing Machinery and Intelligence"

1952–69 Look, Ma, no hands! Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine

1956 Dartmouth meeting: "Artificial Intelligence" adopted

1965 Robinson's complete algorithm for logical reasoning

1966–74 AI discovers computational complexity Neural network research almost disappears

1969–79 Early development of knowledge-based systems

1980–88 Expert systems industry booms

1988–93 Expert systems industry busts: "AI Winter"

1985–95 Neural networks return to popularity

1988– Resurgence of probability; general increase in technical depth "Nouvelle AI": ALife, GAs, soft computing

1995– Agents, agents, everywhere . . .

2003– Human-level AI back on the agenda

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- ~~Play a decent game of table tennis~~
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- ~~Play a decent game of table tennis~~
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- ~~Play a decent game of table tennis~~
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- ~~Play a decent game of table tennis~~
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Swedish in real time

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- ~~Play a decent game of table tennis~~
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Swedish in real time
- Converse successfully with another person for an hour

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- ~~Play a decent game of table tennis~~
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Swedish in real time
- Converse successfully with another person for an hour

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Swedish in real time
- Converse successfully with another person for an hour

© 2021 Pearson Education Ltd.

State of the art

Which of the following can be done at present?

- Play a decent game of table tennis
- Drive safely along a curving mountain road
- Drive safely along Telegraph Avenue
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at Berkeley Bowl
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Swedish in real time
- Converse successfully with another person for an hour

© 2021 Pearson Education Ltd.

According to ChatGPT

Task	Status
Play table tennis	❌ No
Drive mountain road	⚠️ Partial
Drive Telegraph Ave	⚠️ Partial
Buy groceries online	✅ Yes
Buy groceries in-store	❌ No
Play bridge	✅ Yes
Prove new theorem	✅ Yes (narrow domain
Molecular biology research	⚠️ Partial
Write funny story	✅ Yes
Legal advice	⚠️ Partial
Live speech translation	✅ Yes
1-hour conversation	✅ Yes
Complex surgery	⚠️ Partial
Unload any dishwasher	❌ No

Pearson

© 2021 Pearson Education Ltd.

Risks and Benefits of AI

"First solve AI, then use AI to solve everything else." Demis Hassabis, CEO of Google DeepMind

Benefits:

- Decrease repetitive work
- Increase production of goods and services
- Accelerate scientific research (disease cures, climate change and resource shortages solutions)

Risks:

- Lethal autonomous weapons
- Surveillance and persuasion
- Biased decision making
- Impact on employment
- Safety-critical applications
- Cybersecurity threats

© 2021 Pearson Education Ltd.

Risks and Benefits of AI

Development of an artificial superintelligence that surpasses human intelligence may pose a significant risk

Analogous to the "Gorilla problem"
Humans and gorillas evolved from the same species, but humans have more control than other primates.

Thus, we should design AI systems in such a way that they do not end up taking control in the way that Turing suggests they might.

© 2021 Pearson Education Ltd.